



RESEARCH PROGRAM ON
Climate Change,
Agriculture and
Food Security



Principles for Archiving and Sharing

March 2018



Stats4SD

This document was produced in collaboration with [Statistics for Sustainable Development](#) and is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](#)

Introduction

This guide explains some of the benefits of archiving research data. We acknowledge intellectual property and highlight ethical aspects such as anonymity and confidentiality of information providers.

Why should I archive my Data?

In the past, and still today in some cases, researchers were reluctant to share their data. This might be because they fear someone else might get the credit for the work they have done, or perhaps they want more time to carry out their own analyses.

However, most funding bodies are now viewing research data as a public resource and, as a condition of funding, are insisting that data are put into the public domain in a timely fashion. Issues of intellectual property, confidentiality, etc., have been, and are continuing to be addressed.

CGIAR has adopted an Open Access policy with respect to Research Data. This means that all research data, generated as a result of research funded by CGIAR programmes, must, subject to confidentiality of respondents, be deposited in a suitable repository and made publicly available as soon as possible. Details of the CGIAR Open Access policy are available [here](#). A copy of the policy is also available as part of this pack.

Advantages of Sharing Data

There are many advantages to sharing data. These include:

- Encouraging scientific enquiry and debate;
- Promoting innovation and potential new uses of the data;
- Developing new collaborations between researchers;
- Maximising transparency and accountability;
- Encouraging the improvement and validation of research methods;
- Eliminating the need to collect the same data again, thus reducing time and costs;
- Increasing the impact and visibility of the research.

Remember that sharing data is a two-way process; while other researchers are able to make use of your research data, you are also able to make use of data shared by others.

Principles

General principles for archiving data might include:

- Publicly funded research data are a public good which should be made openly available in a timely and responsible manner;
- Sufficient metadata should be recorded and made available to enable others to understand the research and the potential re-use of the data;
- Researchers should have a limited period of privileged use of the data collected to enable them to publish the results of their research – the time might vary but is expected to be no

more than 12 months after data collection or within 6 months of publication, whichever is the sooner;

- All users of the data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.

Intellectual Property

When data are archived, the Intellectual Property or Copyright remains with the researcher(s); it does not transfer to the hosting organisation. For example, the CCAFS Baseline Surveys are archived in Dataverse and hosted at Harvard. The Dataverse Project acts as a publisher in this instance but does not have any rights over the data collections it houses. Users of research data should acknowledge the source of their data, and it is therefore useful for data creators to specify in the archive how they would like to be acknowledged. Dataverse, for example, creates a unique citation for each dataset, and part of the terms and conditions of use is that this citation is used in scholarly references.

The guide on Data Ownership and Authorship includes a short section on Rights and Responsibilities which includes Intellectual Property.

Anonymity and Confidentiality

Before archiving data, you should ensure that the dataset is anonymised – i.e. an individual cannot be identified from their data. Obviously, this would include removing names and addresses of individuals, but there are other things to consider. Anonymising data can be time-consuming, so ensure it is adequately planned for.

Quantitative Data

Techniques for anonymising quantitative data may involve removing or aggregating variables or reducing the precision of a variable.

Remove Direct Identifiers

Direct identifiers include names, addresses, telephone numbers, etc. These are generally not needed for secondary research but are collected for checking purposes or to enable follow-up. These variables can easily be replaced by a code in the data.

Aggregate or reduce the precision of a variable

Examples here might include recording just the year or the year and month of birth rather than the full birth date. Detailed geo-references could be problematic as they could identify individuals. They could be replaced by alternative variables that typify the geographical position; e.g. poverty index, population density, altitude, vegetation type, etc. This would maintain the value of the data without disclosing the exact locations.

Restrict the upper or lower ranges of a continuous variable

If the values for an individual are unusual within the wider group researched, you could collapse unusually large or small values into a single code. For example, a top code of “50 hectares or more” could be applied to land ownership.

Qualitative Data

When anonymising qualitative material such as transcribed interviews, identifiers should not be crudely removed as this can distort the data. Instead we suggest using pseudonyms, replacement terms, or vaguer descriptions. The aim is to achieve a reasonable level of anonymisation whilst maintaining maximum content.

Suggestions for the anonymisation of text include:

- Don't collect personal data unless this is necessary – e.g. don't ask for full names if they can't be used in the data;
- Use pseudonyms or replacements that are consistent across the project – e.g. use the same pseudonyms in publications or follow-up research;
- Use find and replace techniques carefully so that unintended changes are not made, and misspelt words are not missed;
- Identify replacements in text clearly – e.g. by using [brackets];
- Keep original versions of data for use within the research team but don't make them public;
- Create a log of all replacements, aggregations or removals; store the log separately from the anonymised data file. The following table gives an example of such a log file:

INTERVIEW NUMBER	ORIGINAL VALUE	CHANGED TO
1	Age 54	Age range 50-55
1	20 th June	June
1	Cathy (real name)	Jane (pseudonym)
2	Station Hill Primary School	A primary school
2	Rachel	My friend

Anonymising audio-visual data is more difficult, as obscuring faces or altering voices can reduce the usefulness of the data. If confidentiality of audio-visual data is an issue, it is better to obtain the participant's consent to use and share the data unaltered.

Consent

Informed consent is an ethical requirement for most research and must be considered and implemented throughout the research lifecycle from planning through to publication and archiving. Gaining consent must make provision for sharing data.

Researchers should inform participants about how the data they are collecting will be stored, preserved and used in the long-term and how confidentiality will be maintained. It is customary to provide an information sheet to the participants detailing the project and what their involvement will be if they agree to participate. This information sheet should cover the following topics:

- The purpose of the research;
- What is involved in participating;
- Any benefits and/or risks;
- How the data will be used;
- How the data will be stored and used in the future;

- Procedures for maintaining confidentiality;
- Details of the research including the funding source, who is sponsoring the project, and contact details for researchers.

You need to consider the type of data that you will be collecting and whether you intend to follow-up the individual at a later date. Bear in mind that a respondent may be happy to participate initially but may not want to be involved in any follow-ups.

You will need a consent form, and this should allow the participant to clearly respond to each of the following points:

- They have read and understood information about the project;
- They have been given the chance to ask questions;
- They voluntarily agree to participate in the study;
- They understand that they can withdraw at any time without giving a reason;
- They understand that they can refuse to answer one or more of the questions;
- They understand how the data are to be used and archived.

There should also be separate consent sought for use of any audio/visual data such as recordings, videos or photos – some might agree to complete a questionnaire but might not want photos of themselves to be made public.

If your research involves working with children, then consent must be sought from the parent/guardian as well as from the child.

What should I archive?

Of course, when we archive data it is not just the data file itself that we archive. Many data files are of limited use without the accompanying documentation. At the minimum your archive should include:

- The Activity Protocol so others can clearly see the focus of your research;
- The Data Management Plan to show the steps you intended to follow to ensure high quality data;
- The Data Entry system if one has been used;
- The Fieldworker Manual which will detail the procedures used to collect the data;
- A blank copy of the Questionnaire – adding variable names to the questionnaire would be useful for interpreting the data;
- The Data Quality Report which would highlight any problem areas in the data and give suggestions for their use;
- The Metadata Document used to describe the data;
- Etc.

This pack includes a separate checklist of data and documents to submit for archiving.

Preparing for Archiving – Using a DDS

We strongly recommend starting to prepare your archive early in the project lifecycle; if all the preparation is left to the end (as is often the case) then you will find yourself struggling to pull together all the documents and information you need as many of those with the required information will have moved on to other projects. Consider using a Data and Document Storage facility (DDS) – see the separate document in this pack on Data and Document Storage for further information. By using a DDS, you can start to build your archive from the start of the project. When you are ready to archive, it will be much quicker and easier to transfer the files across to the repository.

Summary

There are clear benefits to archiving and sharing research data but there are also responsibilities. You must ensure the confidentiality of your respondents, ensure you have informed consent from respondents and ensure your data have been anonymised.

Think of data sharing as a two-way process – if you are not willing to share your data with others, you cannot expect others to share their data with you.

References

CGIAR Open Access and Data Management Policy -

<https://cgspace.cgiar.org/bitstream/handle/10947/2875/CGIAR%20OA%20Policy%20-%20October%20202013%20-%20Approved%20by%20Consortium%20Board.pdf?sequence=4>

UK Data Archive -

<https://www.ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing/consent-forms>

The Dataverse Project -

<https://dataverse.org/>

Associate Videos

Videos accompanying the original release of the CCAFS Data Management Support Pack in 2013 are available as a playlist on the Statistical Services Centre YouTube Channel at

<https://www.youtube.com/channel/UCs7EU95YMjhvNozJKCD92xQ/playlists>. These videos have not been updated since the original release but are mostly still relevant.

In particular the playlist includes a video on “Archiving and Sharing” which is available at:

<https://www.youtube.com/watch?v=H8sO21P5RBc&list=PLK5PktXR1tmNRaUPsFiYlyhg2lui0xgpi&index=10>